Метод главных компонент в задачах анализа бизнес-процессов

Рубцов А.В., к.э.н., доцент кафедры базовых дисциплин
Лесосибирский педагогический институт – филиал СФУ, Лесосибирск, Россия
Мамаева С.В., к.фил.н., доцент кафедры базовых дисциплин
Лесосибирский педагогический институт – филиал СФУ, Лесосибирск, Россия
Киргизова Е.В., к.п.н., доцент кафедры высшей математики, информатики и естествознания, Лесосибирский педагогический институт – филиал СФУ, Лесосибирск, Россия
Лукин Ю.Л., профессор кафедры базовых дисциплин, Лесосибирский педагогический институт – филиал СФУ, Лесосибирск, Россия
Шмульская Л.С., к.фил.н., доцент кафедры филологии и языковой коммуникации Лесосибирский педагогический институт – филиал СФУ, Лесосибирск, Россия
Мальцева М.В., студентка 4 курса филологического факультета
Лесосибирский педагогический институт – филиал СФУ, Лесосибирск, Россия

Аннотация. Известные способы анализа бизнес-процессов предприятия обладают не только высокой вычислительной сложностью, но и проблемной реализацией. Целью работы является рассмотрение метода главных компонент как эффективного решения в задачах анализа бизнес-процессов предприятия. Предлагается рассмотреть математический инструментарий предлагаемой методики, что позволит решить проблемы, связанные с грамотной визуализацией и моделированием бизнес-процессов без потерь значительного количества информации.

Ключевые слова: метод главных компонент, бизнес-процессы, информационные системы, информационные технологии, размерность данных, эконометрические методы.

The method of the main components in the tasks of the analysis of business processes

Rubtsov A.V., Candidate of Economic Sciences, Associate Professor, Lesosibirskij Pedagogical Institute – branch of Siberian Federal University, Lesosibirsk, Russia

Mamaeva S.V., Candidate of Philological Sciences, Associate Professor, Lesosibirskij Pedagogical Institute – branch of Siberian Federal University, Lesosibirsk, Russia

Kirgizova E.V., Candidate of Pedagogical Sciences, Associate Professor,Lesosibirskij Pedagogical Institute – branch of Siberian Federal University, Lesosibirsk, Russia

Lukin Yu.L., Professor of the Department of Basic Disciplines, Lesosibirskij Pedagogical Institute – branch of Siberian Federal University, Lesosibirsk, Russia

Shmulskaya L.S., Candidate of Philological Sciences, Associate Professor, Lesosibirskij Pedagogical Institute – branch of Siberian Federal University, Lesosibirsk, Russia

Maltseva M.V., 4th year student of the Faculty of Philology, Lesosibirskij Pedagogical Institute-branch of Siberian Federal University, Lesosibirsk, Russia

Annotation. Known methods of analyzing business processes of an enterprise have not only high computational complexity, but also problematic implementation. The purpose of the work is to consider the method of principal components as an effective solution to the problems of analyzing the business processes of an enterprise. It is proposed to consider the mathematical tools of the proposed methodology, which will solve the problems associated with competent visualization and modeling of business processes without losing a significant amount of information.

Keywords: principal component analysis, business processes, information systems, information technologies, data dimension, econometric methods.

Введение

Бизнес-процессы любой организации или предприятия играют решающую роль в измерительно-вычислительных системах. В виду своих свойств обеспечение функционирования бизнес-процессов позволяет минимизировать структурные связи, синхронизировать информационную архитектуру любой организации и способы контроля над ней, минимизировать затраты энергии и упростить этапы моделирования.

С развитием теории управления всё больше возрастает потребность качественного анализа, интерпретации и визуализации данных, получаемых в результате обеспечения функционирования бизнес-процессов. Влияние на обеспечение функционирования в контексте анализа бизнес-процессов любого предприятия в большей степени оказывает повышение производительности того или иного программного обеспечения[1].

Цель настоящего исследования заключается в рассмотрении метода главных компонент как эффективного решения в задачах анализа бизнес-процессов.

Результаты исследования

Метод главных компонент – один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации[2]. Используется в различных задачах для представления набора данных, его визуализации иди сохранения максимума информации о данных в минимальном количестве переменных. Также, метод главных компонент решает проблему мультиколлинеарности – наличия линейной зависимости между объясняющими переменными (факторами) при определении количественных характеристик сложных систем[3].

Суть использования метода главных компонент в анализе бизнес-процессов сводится к возможности уменьшить число переменных, выбрав самые изменчивые из них. А именно, по своей математической сути — это переход к

[1] Егорова Д. К., Тимовкин С. Н. Реализация метода главных компонент в вычислительной среде R // Журнал Огарёв-Online. – 2018, № 1. – С. 1-8.
[2] Сулейманов А.М. Анализ экспериментальных данных методом главных компонент. // Известия Казанского государственного архитектурно-строительного университета. – 2005. – № 2. – С. 1-3.
[3] Тихомиров Н.П., Дорохина Е.Ю. Эконометрика // М.: Изд-во Рос. экон. акад. – 2002. – 640 с.

новым переменным[4]. Каждая главная компонента выражена как линейная комбинация исходных регрессоров. Главные компоненты обладают следующими свойствами:

- выборочная корреляция между двумя новыми переменными всегда равна нулю;
- суммарная дисперсия (разброс) всех исходных регрессоров равен суммарной дисперсии всех главных компонентов.

Метод главных компонент при анализе бизнес-процессов является эффективным по той причине, что большинство факторов, которыми обладают данные любого процесса имеют свойства неопределённости и корреляции между собой[5]. Именно эти характеристики становятся препятствием на пути к качественному анализу, направленного на общее улучшение функционирования любого предприятия или организации.

Предположим, что имеются некие исходные переменные x_1 и x_2. Обратим внимание на то, что они являются центрированными, то есть их среднее значение равно 0. Будем считать x_1 и x_2 исходными регрессорами, то есть – переменными, влияющими на искомую переменную y. Метод главных компонент позволит из регрессоров x_1 и x_2, возможно коррелированных между собой, получить две новых переменных. Полученные главные компоненты будут некоррелированы между собой, то есть не будут взаимосвязаны, а также будут представлять взвешенные исходные переменные. Например, первая главная компонента будет равна:

$$pc_1 = \frac{1}{\sqrt{2}}x_1 + \frac{1}{\sqrt{2}}x_2 \qquad (1)$$

Соответственно, вторая главная компонента примет вид:

$$pc_2 = \frac{1}{2}x_1 - \frac{\sqrt{3}}{2}x_2 \qquad (2)$$

[4] Демешев Б. Б., Малаховская О. А. Картографирование BVAR. Препринт // Высшая школа экономики. М. – 2015. – 37 с.

[5] Зибер И. А., Мороз Г. А. Исследование акустической вариативности S методом главных компонент. // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2019. – Том 17. – № 1. – С. 49-65.

Главные компоненты (pc_1, pc_2), или по-другому, — веса, должны подчиняться следующему условию. Сумма квадратов весов в каждой главной компоненте должна равняться единице.

Итак, метод главных компонент создаёт новые переменные, исходя от первоначальных. При этом цель создания этих новых переменных заключается в выборе максимально изменчивых переменных[6]. Поэтому первая главная компонента pc_1 имеет максимально выборочную дисперсию $sVar$ (pc_1). Затем алгоритм подбирает веса таким образом, что разброс первой главной компоненты pc_1 был максимально возможным.

После формирования первой главной компоненты становится возможным определить вторую. Веса для второй главной переменной (pc_2) алгоритмом подбираются так, чтобы, с одной стороны, - вторая главная компонента не была коррелирована с первой, а с другой, чтобы вторая главная компонента имела бóльшую выборочную дисперсию, максимальную из возможных $sVar$ (pc_2).

Рассмотрим пример для пояснения сущности метода главных компонент в моделировании искусственно созданного бизнес-процесса. Представим, что имеется определённый набор данных, отображающих средний бал отчётности и средний балл производительности отделов Предприятия N – табл. 1.

Таблица 1

Набор данных для анализа[7]

Отчётность (ср. балл)	Производительность (ср. балл)
2	5
4	1
0	3

Необходимо дифференцировать отделы, то есть отобрать параметр, который бы наиболее точно установил различия и неоднородность на

[6] Второва И.А., Качалов О.Б., Плесовских К.Ю. Обработка многомерного сигнала на основе метода главных компонент // Труды НГТУ им. Р.Е. Алексеева. – 2010. –№ 1. – С. 1-6.
[7] Составлено авторами

рассматриваемых уровнях. Проводить дифференциацию по параметру отчётности нецелесообразно, поскольку между средними баллами нет видимого разброса. Данные о производительности отделов представляют больший интерес с этой точки зрения, поскольку разброс оценок значительно выше. Соответственно, первой главной компонентой этой задачи, параметр, который нам и следует преобразовать (pc_1) будет производительность.

Имеющиеся наблюдения (2,4,0) и (5,1,3) представляются нецентрированными переменными соответственно в табл. 2:

Таблица 2

Нецентрированные переменные[8]

a_1	a_2
2	5
4	1
0	3

Перед тем как приступить к поиску главной компоненты pc_1, необходимо центрировать имеющиеся наблюдения, то есть от a_1 и a_2 перейти к $x_1=a_1-\overline{a_1}$ и $x_2=a_2-\overline{a_2}$, где $\overline{a_1}$ и $\overline{a_2}$ – средние арифметические по каждому наблюдению переменной a. Таким образом, обратимся к данным (табл. 3):

Таблица 3

Переход к центрированным переменным[9]

x_1	x_2
0	2
2	–2
–2	0

Переходим к поиску первой главной компоненты pc_1. Так, первая главная компонента – некая комбинация весов x_1 и x_2:

---

[8] Составлено авторами

[9] Составлено авторами

$$pc_1 = \alpha * x_1 + \beta * x_2 \qquad (3)$$

где α, β – числа, сумма квадратов которых, по условию, равна 1; x_1, x_2, pc_1 – вектора.

В ходе вычислений целью представляется подобрать такую пропорцию между x_1 и x_2, чтобы разброс переменной pc_1 был максимален, то есть выборочная дисперсия первой главной компоненты должна быть максимальна:

$$\widehat{Var}(pc_1) \to \genfrac{}{}{0pt}{}{max}{\alpha, \beta} \qquad (4)$$

Поскольку x_1 и x_2 имеют нулевое среднее, максимизация выборочной дисперсии будет равносильна максимизации длины вектора pc_1. Поэтому, целесообразным будет максимизировать длину вектора pc_1 следующим образом:

$$|pc_1|^2 \to \genfrac{}{}{0pt}{}{max}{\alpha, \beta} \qquad (5)$$

$$|pc_1|^2 = (2\beta)^2 + (2\alpha - 2\beta)^2 + (-2\alpha)^2 \qquad (6)$$

$$|pc_1|^2 = 8\alpha^2 + 8\beta^2 - 8\alpha\beta \qquad (7)$$

$$|pc_1|^2 = 8 - 8\alpha\beta, \text{т.к. } \alpha^2 + \beta^2 = 1 \qquad (8)$$

Для того, чтобы найти максимум функции $|pc_1|^2 = 8 - 8\alpha\beta$ при ограничении $\alpha^2 + \beta^2 = 1$, проведём деление на $\alpha^2 + \beta^2$, значение выражения которого равно 1:

$$|pc_1|^2 = 8 - 8\alpha\beta = 8 - \frac{8\alpha\beta}{\alpha^2+\beta^2} = 8 - \frac{8*\frac{\beta}{\alpha}}{1+\frac{\beta^2}{\alpha^2}} \qquad (9)$$

Теперь необходимо свести задачу с ограничением $(\alpha^2 + \beta^2 = 1)$ к задаче без ограничения с помощью замены переменной: $\frac{\beta}{\alpha} = t$. Таким образом, получаем следующее выражение, которое затем необходимо привести к максимуму, подобрав наилучшую пропорцию между α и β:

$$|pc_1|^2 = 8 - \frac{8t}{1+t^2} \to \frac{max}{t} \tag{10}$$

Пусть (10) равно $Q(t)$, тогда $Q'(t) = 8 - 8t^2 = 0$. Получаем две точки экстремумов $t_1 = 1$, $t_2 = -1$. Соответственно, точкой максимума будет являться t_2.

В ходе проведённых вычислений было выявлено, что для максимизации разброса первой главной компоненты (4) необходимо взять отношение $\frac{\beta}{\alpha} = 1$ и уже ранее установленное ограничение $\alpha^2 + \beta^2 = 1$. Соответственно, получены два решения:

$$1.\ \alpha = \frac{1}{\sqrt{2}}; \qquad 2.\ \alpha = -\frac{1}{\sqrt{2}};$$

$$\beta = -\frac{1}{\sqrt{2}} \qquad \beta = \frac{1}{\sqrt{2}}$$

Оба решения являются верными в рассматриваемом случае и удовлетворяют вышеописанным условиям. Для нахождения искомой первой главной компоненты pc_1 мы выберем решение 1. Получаем:

$$pc_1 = \frac{1}{\sqrt{2}}x_1 - \frac{1}{\sqrt{2}}x_2 \tag{11}$$

Поскольку нам известны значения x_1 и x_2 из таблицы 3, становится возможным посчитать pc_1 для каждого из трёх наблюдений:

pc_1 (для 0 и 2) = $-\sqrt{2}$;

pc_1 (для 2 и -2) = $2\sqrt{2}$;

pc_1 (для -2 и 0) = $-\sqrt{2}$.

Таким образом, мы получили новую искусственную переменную pc_1, которая линейно выражается через значения центрированных исходных переменных x_1 и x_2 и имеет максимальную дисперсию. Именно это позволяет использовать данную переменную для дальнейшего анализа, её грамотной визуализации и моделирования без потерь значительного количества информации

Заключение

Вышеописанное подтверждает ранее высказанную идею о важности характеристик бизнес-процесса при анализе. Метод главных компонент при выполнении данной задачи позволяет учесть основные условия, характерные для того или иного бизнес-процесса, провести преобразования переменных, которые в дальнейшем будут пригодны для регрессионного анализа и последующего за ним выявления особенностей при моделировании не только рассматриваемого бизнес-процесса, но и сложной динамической составляющей – модели процессов в целом. Таким образом, использование метода главных компонент в задачах анализа бизнес-процессов представляет перспективу развития как информационных технологий в целом, так и применения эконометрических методов в теории управления.

Библиографический список

1. Второва И.А., Качалов О.Б., Плесовских К.Ю. Обработка многомерного сигнала на основе метода главных компонент // Труды НГТУ им. Р.Е. Алексеева. – 2010. – № 1. – С. 4.
2. Демешев Б.Б., Малаховская О.А. Картографирование BVAR. Препринт // Высшая школа экономики. М. – 2015. – С.18.
3. Егорова Д. К., Тимовкин С. Н. Реализация метода главных компонент в вычислительной среде R // Журнал Огарёв-Online. – 2018. – № 1. – С. 6.
4. Зибер И.А., Мороз Г.А. Исследование акустической вариативности S методом главных компонент. // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2019. – Том 17. – № 1. – С. 51.

5. Сулейманов А.М. Анализ экспериментальных данных методом главных компонент. // Известия Казанского государственного архитектурно-строительного университета. – 2005. – № 2. – С. 2.

6. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика // М.: Изд-во Рос. экон. акад. – 2002. – С. 236.

References

1. Vtorova I.A., Kachalova O.B., Plesovsky K.Yu. Processing of a multidimensional signal based on the method of main components // Trudygtu NSTU named after R. I. Alekseev. – 2010. – № 1. – P. 4.

2. Demeshev B.B., Malakhovskaya O.A. Cartografiya BVAR. Overcoming / Higher School of Economics, Moscow. – 2015. – P. 18.

3. Egorova D.K. Timovkin S.N. Implementation of the main component method in the computing environment R // Journal Ogarev-Online. –2018. – № 1 – P. 6.

4. Sieber I.A., Moroz D. A. Study of acoustic variation s method of the main component. // Bulletin of the Novosibirsk Pre-trial University. Series: Linguistics and intercultural communication. – 2019. – Volume 17. – № 1. – P. 51.

5. Suleymanov A.M. Analysis of expert data of the main component methods. // News of the Kazan Hospital of the University of Architecture and Civil Engineering. – 2005. – № 2. – P. 2.

6. Tikhomirov N.P., Dorohina O.Yu. Econometrica // Moscow: Publishing House of Ross. ekon. akad. – 2002. – P. 236.